POWER SUPPLY NOISE ANALYSIS FOR 3D ICS

USING THROUGH-SILICON-VIAS

A Thesis
Presented to
The Academic Faculty

by

HEMANT SANE

In Partial Fulfillment
of the Requirements for the Degree
MASTERS in the
School of ELECTRICAL AND COMPUTER ENGINEERING

Georgia Institute of Technology
MAY 2010

POWER SUPPLY NOISE ANALYSIS FOR 3D ICS

USING THROUGH-SILICON-VIAS

Approved by:

Dr. Sung Kyu Lim, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Hsein Hsin Lee
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Gabriel Loh
College of Computing
Georgia Institute of Technology

Date Approved: January 12, 2010

ACKNOWLEDGEMENTS

I wish to thank all members of GTCAD lab, whose assistance was essential in completing my work.

TABLE OF CONTENTS

		Page
ACKNOWLEDGEMENTS		iii
LIST OF TABLES		v
LIST OF FIGURES		vi
LIST OF SYMBOLS AND ABBREVIATIONS		vii
SUMMARY		viii
CHAPTER		
1	INTRODUCTION	1
2	NEED OF RESEARCH	5
3	P/G OPTIMIZATION FOR NOISE MINIMIZATION	7
	RLC MODEL	10
	PROBLEM FORMULATION	11
	VOTLAGE DROP ESTIMATION	12
	OPTIMIZATION ALGORITHM	13
	FULL CHIP SIMULATOR	15
	RESULTS FOR P/G NETWORK OPTIMIZATION	16
4	3D NOISE ANALYSIS USING 2D TOOLS	22
	FLOW FOR 3D IR DROP ANALYSIS	22
	RESULTS FOR 3D NOISE ANALYSIS – 3DMAPS	27
	CADENCE LIMITATIONS FOR 3D IR DROP ANALYSIS	30
5	CONCLUSION AND FUTURE WORK	32
REFERENCES		33

LIST OF TABLES

	Page
Table 1: IR Drop	19
Table 2: Dynamic Noise	19
Table 3: Time taken to run simulation (full chip v/s unit model) for 1-5 die case	20

LIST OF FIGURES

Page

Figure 1: Types of die stacking 2

Figure 2: Types of TSVs 3

Figure 3: Supply bounce and ground bounce 4

Figure 4: 3D Power distribution network 9

Figure 5: Grid tile for optimization 11

Figure 6: Voltage drop for 5 cores v/s time 16

Figure 7: Size distribution for uniform, optimized, and congestion-aware cases 17

Figure 8: % Degradation in IR drop over a maximum sized network as function of number of core dies 18

Figure 9: % Degradation in dynamic noise over a maximum sized network as function of number of core dies 18

Figure 10: Full chip simulation results 21

Figure 11: 3D Process description file for face-to-back stacking 23

Figure 12: The 3DMAPS core die 28

Figure 13: The 3DMAPS memory die 29

LIST OF SYMBOLS AND ABBREVIATIONS

3D	three dimensional design/object
n	number of thin metal wires in single grid tile in one direction
r	ratio of widths of thick metal wire to thin metal wire
w_{thin}	width of a single thin wire
TSV	Through-Silicon-Via
P	Power-to-power bump TSV pitch

SUMMARY

3D design is being recognized widely as the next BIG thing in system integration. However, design and analysis tools for 3D are still in infancy stage. Power supply noise analysis is one of the critical aspects of a design. Hence, the area of noise analysis for 3D designs is a key area for future development. The following research presents a new parasitic RLC modeling technique for 3D chips containing TSVs as well as a novel optimization algorithm for power-ground network of a 3D chip with the aim of minimizing noise in the network. The following work also looks into an existing commercial IR drop analysis tool and presents a way to modify it with the aim of handling 3D designs containing TSVs.

CHAPTER 1

INTRODUCTION

Silicon industry today runs according to a popular form of Moore's law, which states that "transistor-count on a single wafer shall double every 18 months". Conventional practice for digital devices is to reduce the minimum dimension allowed. Thus, progress is accompanied by continuously moving from one technology node to the next. Today, we are close to the saturation point. Gate sizes can be reduced only so much further. Instead, the latest buzzword is 3D, or three-dimensional design. Design today is shifting gears from basic two-dimensions or 2D, to 3D. Silicon area is at a premium, and customers are demanding more value for their money. Plenty of time and money is being put by university labs as well as industry R&D folks to identify possible ways to stack dies on top of another. Three dimensional circuit technologies implement multiple tiers of active dies stacked above each other. This is being heralded as the means to continue along the path of increased integration along the Moore's law curve. It has the potential to increase packing density and reduce chip area significantly, in comparison with today's 2D ICs[1]. However, even though process technology nodes have been shrinking at a fast rate, these advancements have not been uniform. The disparity between transistor and wire scaling is quite prominent leading to obstruction in further performance scaling. 3D integration seems to provide the answer to this issue too.

Two dies can be stacked on top of one another in three ways as shown in figure 1[2]. Face-to-back stacking is the most popular, though in some instances, face-to-face

stacking is also used. For face-to-back stacking, a vertical cross-section multi-die stacked chip shows alternate layers of silicon bulk and metal, as seen in figure below. Back-to-back stacking is the third type of stacking.

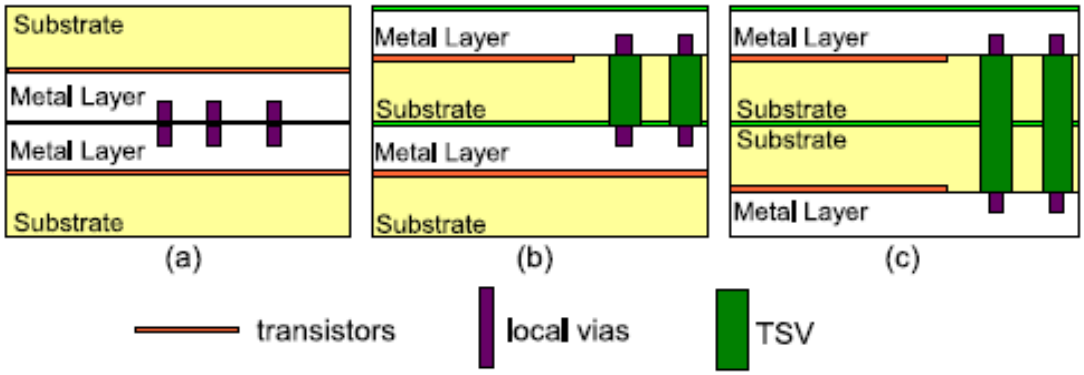


Figure 1: Types of die stacking

(a) Face-to-face stacking (b) Face-to-back stacking (c) Back-to-back stacking

A through-silicon-via (TSV) is a vertical electrical connection passing completely through a silicon wafer / die. Through-silicon-vias (TSVs) are used to connect a gate on one die to another on a die above/below the current one. TSVs are actually holes drilled straight through the bulk, and coated with copper from inside. This drilling can occur either before metallization, or after metallization. Accordingly, they are classified into two main categories, via-first and via-last respectively, as shown in figure 2[2].

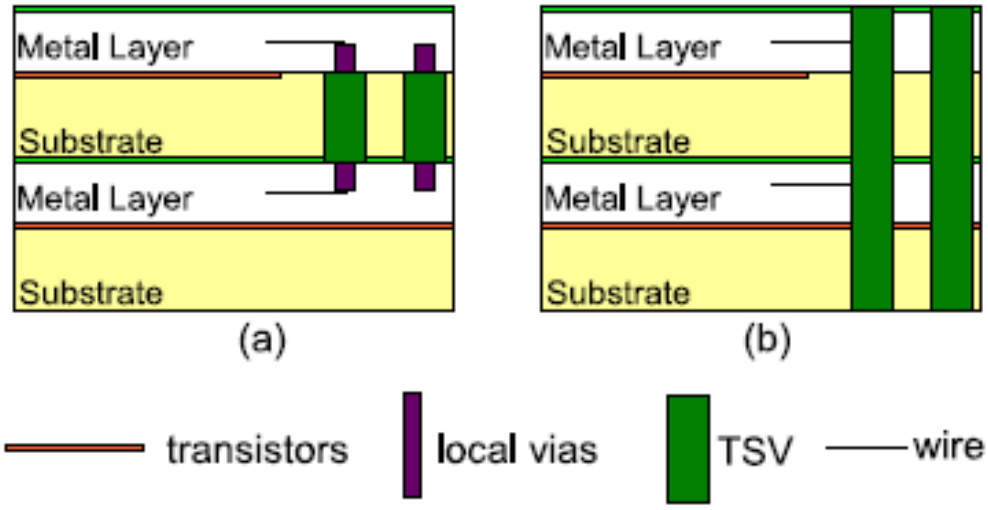


Figure 2: Types of TSVs

(a) via-first TSVs (b) via-last TSVs

Since TSVs are electrical connections made using copper, TSV parasitics do exist and have to be taken into account during all design and analysis steps. One area affected by TSV parasitics is the noise analysis of the design. TSV parasitics add to existing parasitics in the power-ground network, and cause additional voltage drop.

It is important that supply voltage reaches the circuits as efficiently as possible, minimizing any voltage drop from external source to the circuits. Similarly, care must be taken to minimize any ground bounce. Figure 3 shows the reason supply and ground bounce exist.

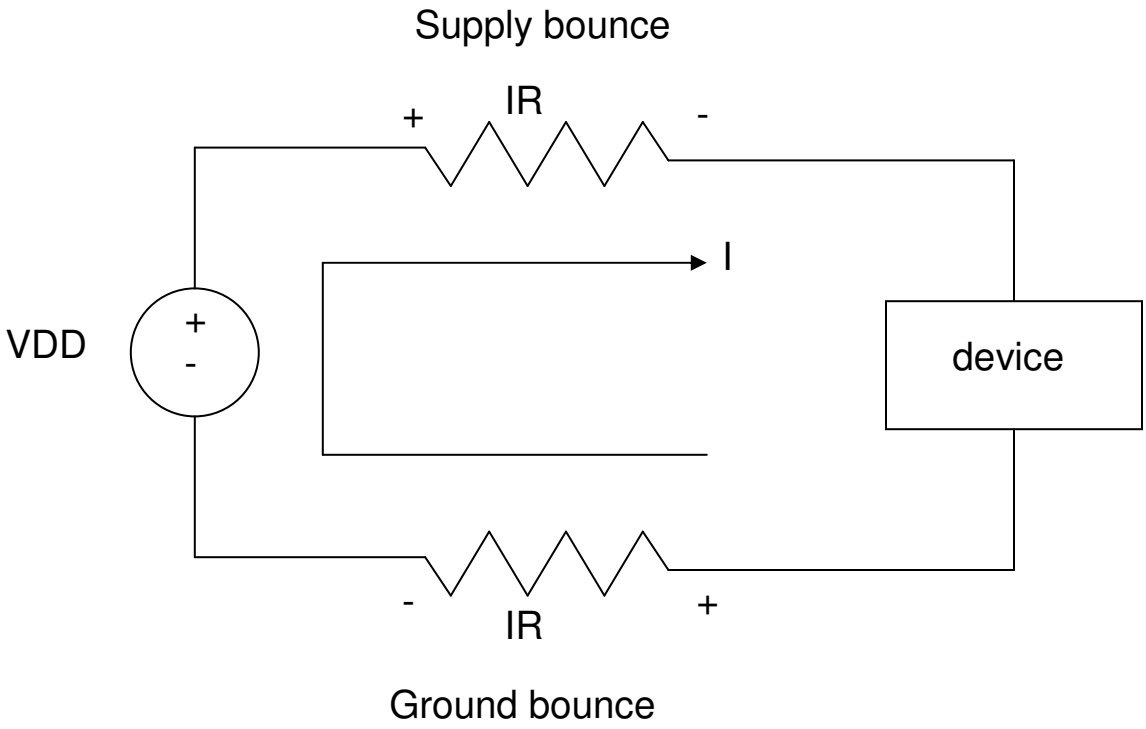


Figure 3: Supply bounce and ground bounce

Supply noise (supply and ground bounce) can be classified broadly into two categories. Static noise is basically just the IR drop (shown in figure 3). Dynamic noise is the total noise and covers effect of inductance and capacitance (parasitics) on supply voltage and ground voltage values. Dynamic noise is seen as a fluctuation over actual value of the supply line. Worst case drop occurs at the first trough on the fluctuations, and is called as 'first droop'. Different techniques are used to increase reliability of power grids and control power grid noise today. They include wire widening, grid topology optimization and decoupling capacitor insertion[3]. Noise analysis tools to measure supply network noise exist commercially. But currently they can handle only 2D designs. They need to be expanded to apply them for 3D designs.

CHAPTER 2

NEED OF RESEARCH

Noise analysis results are one of the important criteria for judging the quality of any design. This remains true even for 3D designs. In fact, when it comes to 3D design, designers have more issues to take care of to achieve supply voltage levels within noise margin. A k-tier 3D chip that stacks k similar chips could use k times as much current as used by a single chip with same footprint. And since packaging technology has not been able to advance at an equal pace as process technology, with a similar number of pins in package, current per pin can be up to k times higher in the 3D chip[4]. Also, TSVs are now a part of the power distribution network, and they have their own set of parasitics. This will add to the overall resistance of the power distribution network, and hence will cause a rise in the supply bounce and ground bounce. At the same time, noise margins are becoming more stringent with reduced VDD levels that are required for smaller process technology nodes. Hence, noise analysis for 3D is an important area that needs looking into.

We will be considering two topics

1. Power grid optimization using parasitic RLC model for 3D chips containing through-silicon-vias – theoretical study
2. Expanding 2D tools to analyze 3D chips – practical study

1) One widely used method for power-ground analysis is simulation of power-ground network using a spice model of the entire network. Complexity of the model increases by a large extent for 3D chips, and computational time is directly related to the model complexity. Hence there is a need for a new approach to creating the spice model that will reduce computational time without compromising on accuracy of results.

2) Existing design automation tools are designed to handle 2D designs, but not 3D designs. There is a need for tools that can design and analyze 3D designs. But we are caught in a paradox here. To design 3D tools, we need to know how 3D designs work; but we need 3D tools to be able to make and analyze 3D designs. Hence, the best way out at this stage is to improvise on existing material. Noise analysis is one of the key ways of judging quality of design, and so getting a 3D noise analysis tool is necessary.

CHAPTER 3

P/G GRID OPTIMIZATION FOR NOISE MINIMIZATION

Power/ground network for a 3D design is much more complex as compared to a 2D design. In fact, it contains multiple 2D power/ground networks connected using TSVs. Hence, it is time consuming to do noise analysis on 3D design by performing spice simulations on its complete netlist.

An alternate method is proposed here. Instead of performing spice simulations on the entire netlist, divide the 3D design into 3D unit models. Total time is equal to summation of time required to perform noise analysis on each single 3D unit model, which is much less than the time required for performing noise analysis on entire 3D design as a single unit. At the same time, loss of accuracy is negligible.

Maximum voltage drop is shown to be a function of parameters from the power grid model. By optimizing the values of these parameters, overall voltage drop (noise) can be minimized.

For this study, Intel Penryn architecture is chosen as the baseline design. One core die contains two twin-core processors giving four cores per die. The system contains up to five of such core dies stacked in vertical fashion using face-to-back stacking. The dies are connected using TSVs. 3D stacking technology requires die thinning to very small thickness. However, 3D stacking gives rise to thermal issues of a more serious level than those in 2D technology. Power density values offered will be larger than what standard air-cooled heatsinks can cope with. Recently, work is being carried out on micro-fluidic

channels implemented onto the backs of the 3D stacked dies in order to use fluids for cooling[5]. This study assumes use of micro-fluidic heatsinks to dissipate heat. In order to accommodate size and mechanical stresses of micro-fluidic channels, die thickness is kept to $150\mu m$. TSV height is kept equal to the die thickness. The power grid for this system comprises of TSVs, thick wires and thin wires. Uniform distribution of decoupling capacitors is assumed.

For multi-level stacked chips with number of dies greater than two, power/ground TSVs are stacked into vertical columns. These power/ground TSVs are connected to power/ground balls (bumps) from the ball grid array. The TSVs are assumed to be made of copper. A conservative estimation of resistivity, $21n\Omega.m$, is made to account for any thermal effects. At each die level, adjacent power TSVs are connected horizontally by thick wires (width = 10 um) forming a coarse grid. Similar grid exists for ground TSVs. Within this grid is a finer mesh of thin wires (width = 5um). The pitch for the TSVs is 400 um. The ground grid is offset from the power grid by half the TSV pitch i.e. 200 um. Thus, at each level, there exists a 2D network of thick and thin wires. Figure 4 shows the general topology of 3D power distribution network[6].

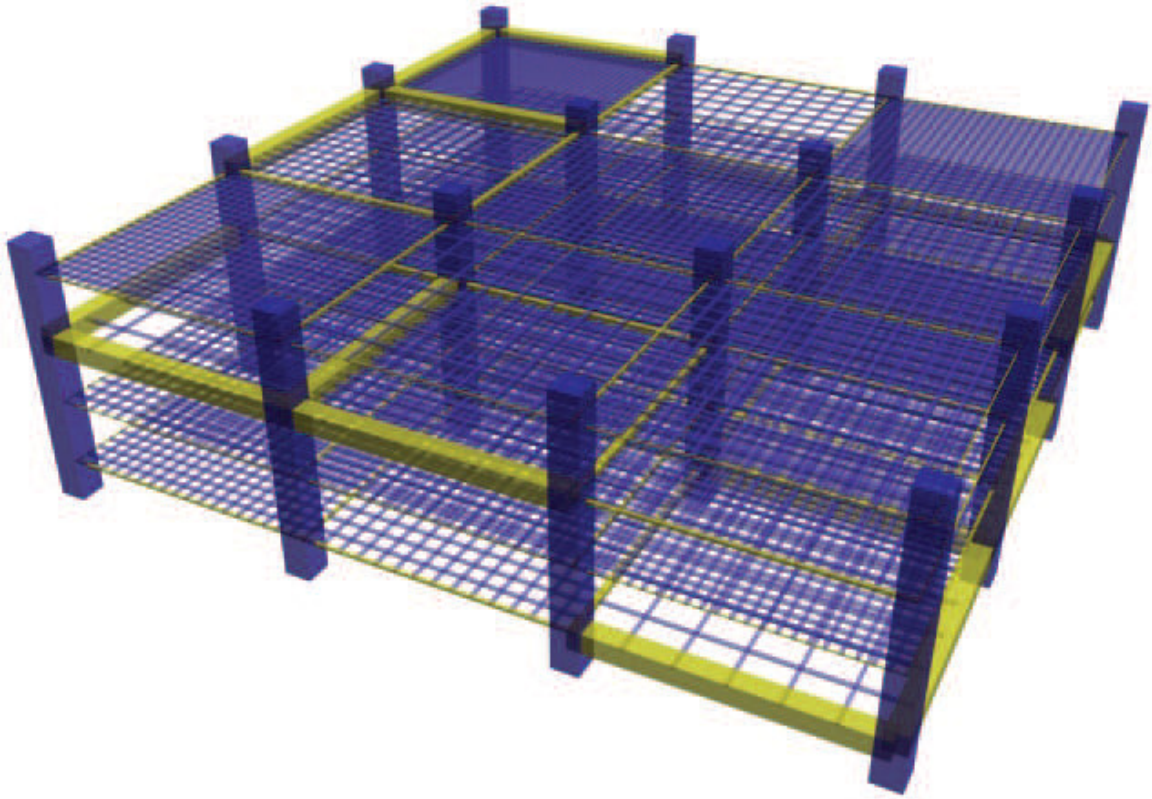


Figure 4: 3D Power distribution network

The TSVs have a height of 150 μm. TSV pitch is kept much larger than its length. This ensures mutual-inductance of neighboring TSVs is dominated by self-inductance of the TSVs, and can be safely ignored. This is important since TSV inductance adds to the already significant inductance value of the package, and this in turn results in larger low-frequency “first droop” noise. Also, uniform current density is assumed so that no high-frequency oscillation is present to cause skin effect to become dominant.

A related study was performed earlier that examined large scale stacking of 3D ICs[6]. This study draws from it and focuses on power-ground network optimization for 3D ICs.

RLC MODEL

A flip-chip package with ball grid array chip connections (bumps) is assumed. A further assumption is that TSVs and package bumps contain parasitic resistance, capacitance and inductance. The 2D network at each die level is assumed to be purely resistive. A capacitor (representing a decoupling capacitor) and a current source (modeling current demand of the transistors) are connected between corresponding nodes of power and ground grid. The current sources values are ramped up from 0 to value equal to current demand for that region of the floorplan. Current sources are also used to model current demand of other dies such that current flowing through any TSV is equal to current demand on the die being examined, added to current demand of all dies above it. Experimentally it can be shown that spice simulations results for the case when lower dies are modeled using current source are identical to the case when lower dies are modeled using parasitic elements. Decoupling capacitor is connected to each node, with its value obtained by assuming uniform distribution of decoupling capacitors over entire die.

The package model is designed as a hybrid of lumped and distributed models. The C4 package bumps have impedance that is clubbed together with package distribution impedance. The total impedance is modeled using a 5 mΩ resistor, a 500 pH inductor and a parallel 30 fF capacitor from the bottom of the power/ground TSVs to ground[7]. TSV parasitics have been extracted using[8]. TSV inductance is assumed to be self-inductance only and is 55pH. TSV capacitance is taken to be 48fF.

PROBLEM FORMULATION

The objective is to reduce the worst case dynamic noise in 3D stacked ICs. As defined earlier, power/ground network is described as a 2-level hierarchical mesh. Higher level contains thick wires connecting neighboring power pads/TSVs. Lower level contains thin wires performing local distribution. ‘n’ denotes number of thin wires between neighboring power pads/TSVs, while ‘r’ denotes ratio of thick wire width to thin wire width. ‘i’ is tile number taken from universal set G containing all grid tiles of die under consideration. Figure 5 shows a grid tile for optimization.

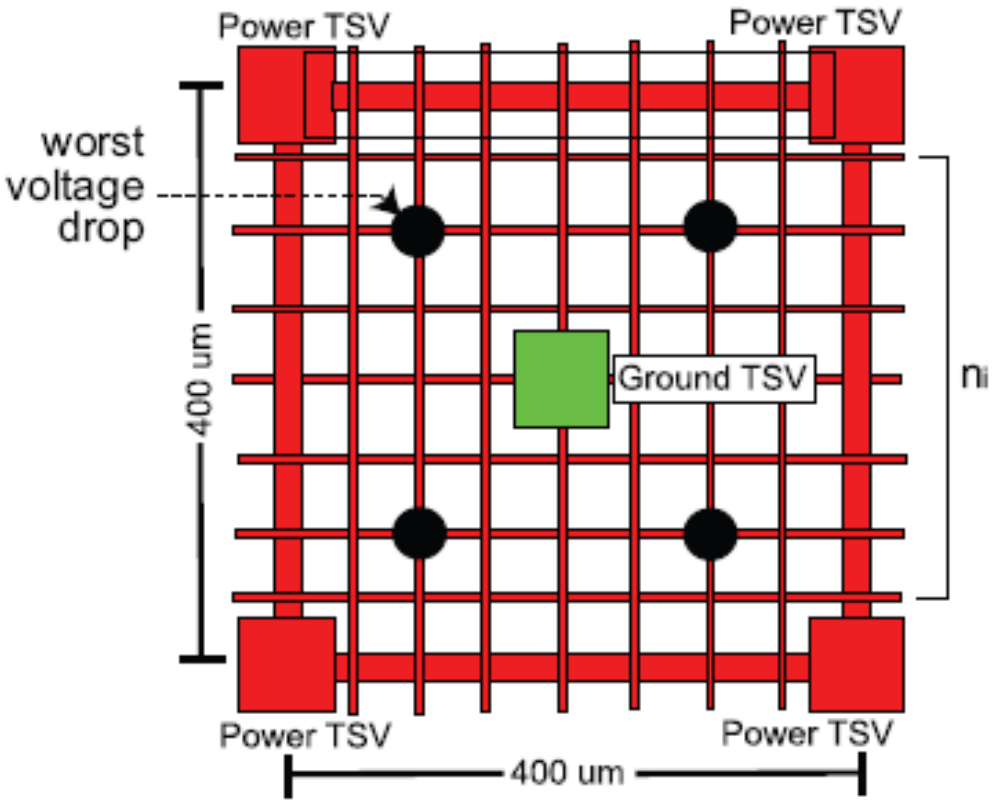


Figure 5: grid tile for optimization

Dynamic noise minimization problem is defined as follows:

Objective Function: Minimize: MAX $[V_{drop}(n_i, r_i)]$

Constraints:

i) Maximum area: this constraint ensures area of resulting power grid is not too large. Maximum allowable area (A_{max}) is kept equal to that resulting from a uniform grid.

$$\sum_i (2P \cdot w_{thin} \cdot n_i + 2P \cdot w_{thin} \cdot r_i) \leq A_{max}$$

ii) Wire width: Maximum wire width is defined by manufacturing constraints. This constraint ensures the limit is not crossed.

$$r_i \leq r_{max}$$

iii) Congestion: By having a maximum number (C_i) for thin wires that can be put in a grid tile while still maintaining routability of the tile, congestion can be limited. This constraint is not present in basic optimization algorithm. Addition of this constraint converts the basic optimization algorithm to congestion-aware optimization algorithm.

$$n_i \leq C_i$$

VOLTAGE DROP ESTIMATION

It is observed that correlation between grid tiles separated by more than one power-to-power pitch is very low. Also, voltage drop for a tile in one tier is not affected by *n* and *r* values for same tile in other tiers. This allows simulation of each grid tile of the power map and its surrounding area individually, using SPICE. Voltage drop numbers for that tile then are stored in a table indexed by *n* and *r*; such that n_i is an odd integer value and

$n_i \in [3, 21]$ and r_i belongs to a set of 16 evenly spaced values between [0.5, 3.5] . System under study has die length and width of 5.6μm. A TSV pitch length of 400 μm gives a total of 196 tiles.

For a single value of *r*, maximum voltage drop in grid is obtained for different values of *n* using Hspice. For each value of *r*, a graph of V_{drop} v/s *n* is obtained. This graph is curve fitted to a 3rd degree polynomial.

$$\text{Equation: } V(n, r) = a + b/n + c/n^2 + d/n^3$$

TSV drop is modeled as $V_{TSV} = I_{TSV}.R_{TSV}$

$V_{total} = V_{tile} + V_{TSV}$

$= \text{beta } (a + b/n + c/n^2 + d/n^3) + I_{TSV}.R_{TSV}$

where beta is a multiplying factor.

OPTIMIZATION ALGORITHM

The processor power map is based on the Intel Penryn architecture. From a publicly released die photo of the Penryn architecture (45nm) a floorplan was generated. The total power dissipation (54 watts for dual-core version) was divided into component stages of logical pipeline for the two processors giving a power map for one core-die.

Power map of a die contains a power dissipation number for each module in the floorplan. This power dissipation number is based only on the gates on the die itself, assuming stand-alone die. This power dissipation number is converted to a power density number using area of each module. Then this floorplan is divided into a grid of tiles using power TSVs as vertices. Each grid tile is assigned a power density number equal to maximum power density of all blocks that occupy the said tile, forming a modified power

map. When the die is placed in a 3D stack, power dissipation values will change due to current drawn by gates in other dies. Taking this effect into account, the modified power maps for individual dies are converted into a single 3D current density map keeping grid tile size identical to that of the modified power map.

As a result, each tile is assigned current density equal to that of the block having largest current density covering that tile. Thus, every tile has a current density number and congestion constraints, which are represented as a set of n_i and r_i values. Results are compressed such that all tiles with same current density value and same congestion constraints are represented by the same set of n_i and r_i values. Since it is a nonlinear optimization, runtime increases exponentially. Hence, compression helps in reducing the runtime significantly (days reduced to minutes). Also, it does not have any negative effect on quality of the result.

This algorithm is implemented as a MATLAB script using the non-linear optimizer function *fmincon*, which uses a sequential quadratic programming (SQP) method. The function provides optimum values for n_i and r_i, which are continuous variables. Final values are obtained by rounding. The optimizer calculates the gradient using perturbations of a single variable. If two or more tiles have the same voltage drop then the return value of the Max function (objective function) will be unchanged by negative single variable perturbations. Therefore we change the objective function to

$$\sum_i (V_{drop}(n_i, r_i) \cdot MAX(V_{drop}))$$

By supplying a limit for maximum voltage drop, and a starting value of *r* to the optimizer, the optimizer can print out optimum values of *n* and *r* under given design

constraints. The limit for maximum voltage drop can be called as a noise threshold, usually at 10% of VDD.

In 3D, the underlying assumption is that worst case drop will always occurs for die furthest from bumps. However, this algorithm optimizes power-ground network for all dies, to ensure voltage levels for the die furthest away from the bumps is as close to ideal as possible.

FULL CHIP SIMULATOR

Simulation of power grids is a major issue in itself because of the number of nodes they contain. Most commercial tools do not have the capability to handle a netlist from a 5 die 3D IC. Hence, a custom circuit simulator was used[6]. It is based on Modified Nodal Analysis[9]. This custom circuit simulator builds a uniform P/G network for given design with parameters like number of dies, dimension of each die, TSV dimensions, parasitic values, and few others. This network is actually a parasitic model comprising of resistors (R), capacitors (C), and inductors (L) along with current sources to model current demand. This network can be analyzed by running SPICE simulations on it. The full chip simulator is used as a baseline here to compare accuracy of results obtained from the optimization algorithm.

RESULTS FOR P/G NETWORK OPTIMIZATION

Total voltage drop (supply bounce and ground bounce) was observed for each of the five dies over a period of time. Figure 6 shows variation in supply drop for each of the five cores with time. Each line in the graph represents the noisiest point for each tier.

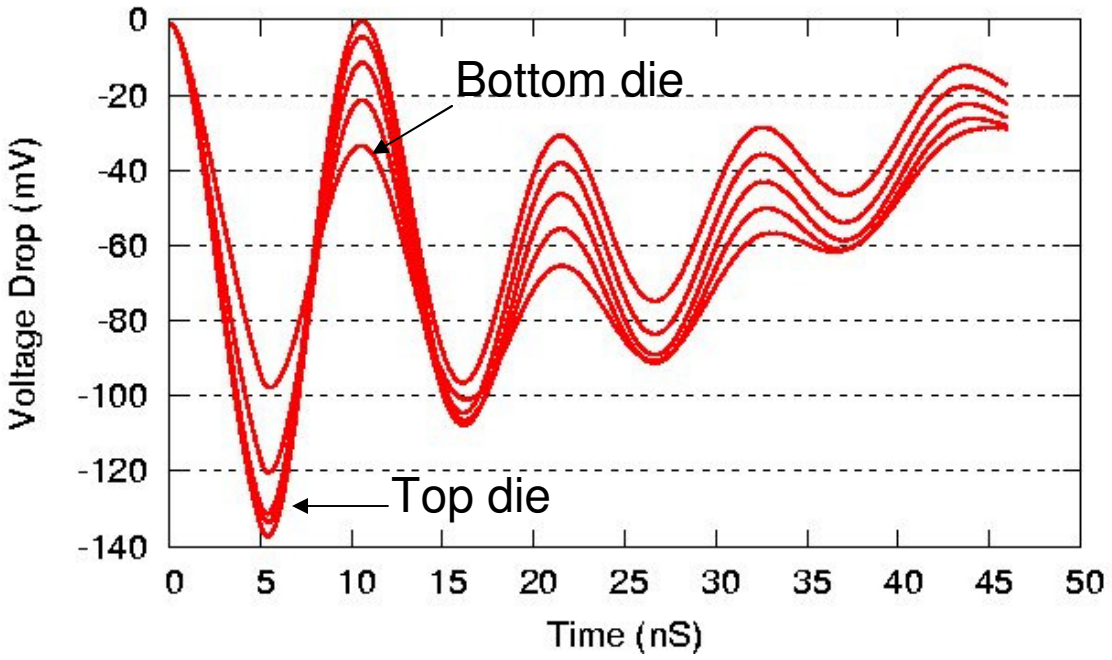


Figure 6: Voltage drop of 5 cores v/s time

Maximum drop is observed at first trough on the wave, which is called as “first droop”. These fluctuations occur due to parasitic inductance and capacitance seen by the supply. As seen here, top die, which is furthest away from the power bumps, shows the maximum noise levels.

Figure 7 shows a comparison of histograms for distribution of *n* for the uniform, optimized and congestion-aware cases.

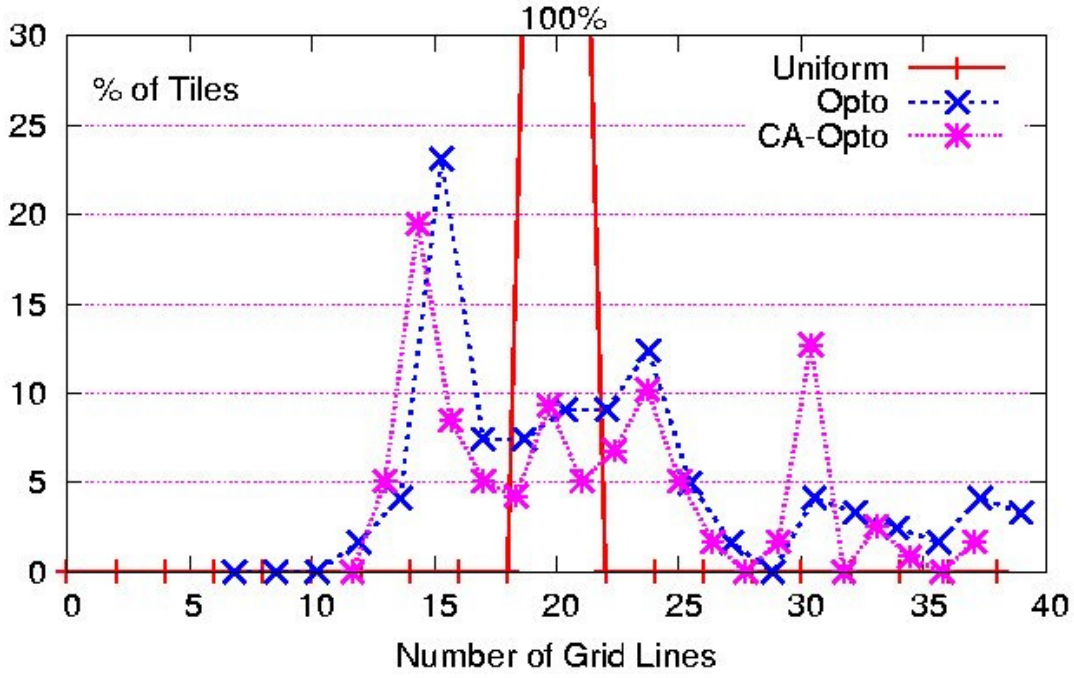


Figure 7: Size distribution for uniform, optimized, and congestion-aware cases

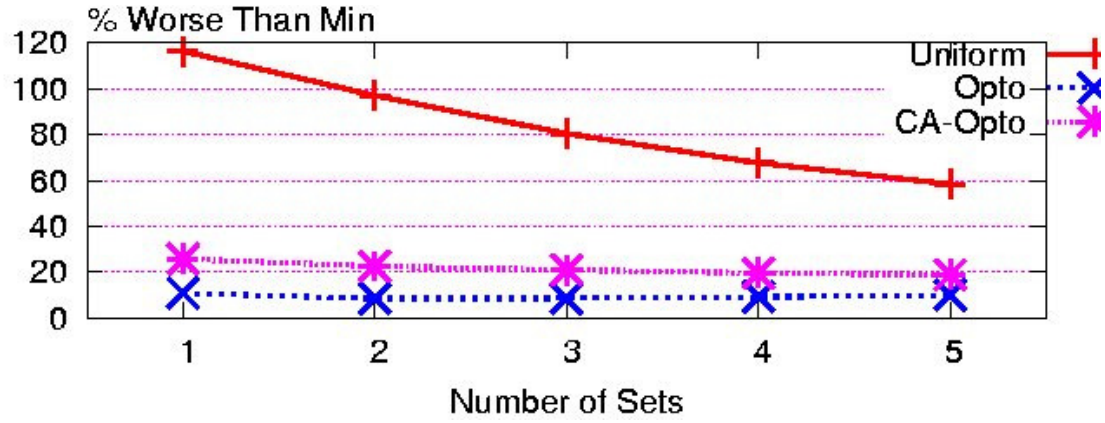
Three cases were considered to describe benefits of the algorithm.

Case 'Uniform': All tiles have uniform values for *n* and *r*.

Case 'Opto': Values of n and r for each tile are determined by the optimization algorithm. Congestion constraint is not added to the problem.

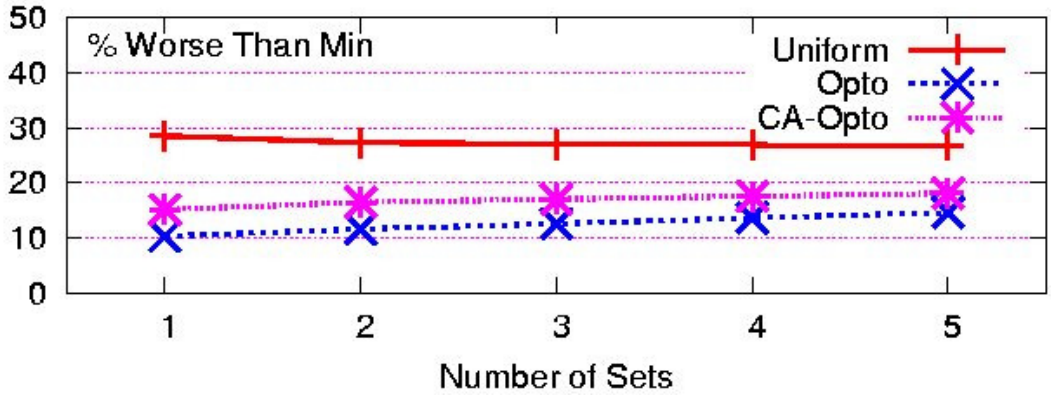
Case 'CA-Opto': Values of n and r for each tile are determined by the optimization algorithm with congestion constraint added to the maximization problem.

A power-grid where all space is utilized for power distribution is used as a baseline. The results of the optimization algorithm are compared with this theoretical minimum noise level, or 'Min', obtained from the baseline power-grid. Figures 8 and 9 show the %age degradation in IR-Drop and dynamic noise for the three cases.



IR Drop

Figure 8: % Degradation in IR drop over a maximum sized network as function of number of core dies



Dynamic Noise

Figure 9: % Degradation in dynamic noise over a maximum sized network as function of number of core dies

Tables 1 and 2 give IR drop and dynamic noise values for all five dies; for each of the three cases, along with the theoretical minimum.

Table 1: IR Drop

	Die 1	Die 2	Die 3	Die 4	Die 5
Min	16.1912mV	19.6506mV	24.2232 mV	29.7448 mV	36.0725 mV
Uniform	35.0436mV	38.7192mV	43.7024 mV	49.8722 mV	57.1288 mV
Opto	17.9479mV	21.3071mV	26.3039 mV	32.4783 mV	39.6725 mV
CA-Opto	20.3669mV	24.0857mV	29.2955 mV	35.5659 mV	42.8649 mV

Table 2: Dynamic Noise

	Die 1	Die 2	Die 3	Die 4	Die 5
Min	67.4928mV	91.4861mV	109.935 mV	125.965 mV	140.709 mV
Uniform	86.7742mV	116.49mV	139.665 mV	159.889 mV	178.453 mV
Opto	74.4378mV	102.1259mV	123.7538mV	143.1466mV	161.2525mV
CA-Opto	77.7179mV	106.5904mV	128.701mV	148.1222mV	166.1772mV

These graphs and tables show the effectiveness of algorithm, with wiring resources as constraint. Effect is more prominently seen in IR drop (static) than in dynamic noise. This is because dynamic more dependent on decoupling capacitance distribution than wiring resources.

Table 3 gives comparison between the times required to run simulations using the optimization algorithm against the custom full chip simulator, for one-die case to five-die case.

Table 3: Time taken to run simulation (full chip v/s unit model) for 1-5 die case.

Time	Simulation using single model for full chip	Simulation of full chip using unit models
1 die	2 hours	10 min
2 die	4.5 hours	17 min
3 die	13 hours	22 min
4 die	28 hours	28 min
5 die	56 hours	40 min

* These simulations were performed on a machine with reasonable processing capability.

As seen from the table, the algorithm provides time saving by at least one order of magnitude. Also, time savings increase as design size increases.

Figure 10 gives full chip simulation results for voltage drop per grid tile for five die case. These plots are for the die furthest away from the power bumps. As shown in figure 6, this die shows the worst case noise levels in a five die chip.

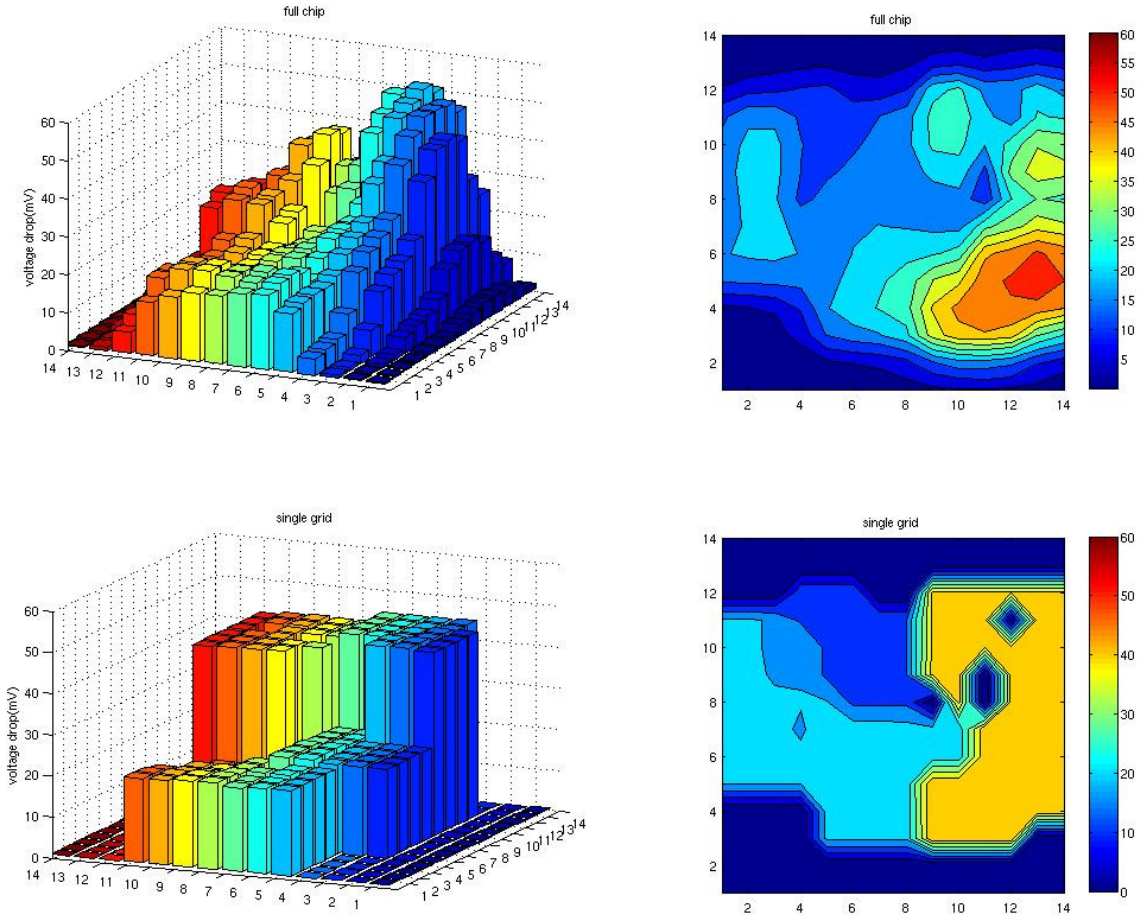


Figure 10: Full chip simulation results

The top plots in figure 10 show maximum IR drop per grid-tile using nodal simulator, while the bottom plots show maximum IR drop per grid-tile obtained using optimization algorithm and unit models for P/G network. There is a marginal loss of accuracy. This loss is attributed to having conservative algorithm constraints.

CHAPTER 4

3D NOISE ANALYSIS USING 2D TOOLS

3D noise analysis is demonstrated using Cadence® tools. Cadence® Encounter® family of products provides an integrated solution for RTL-to-GDSII design flow[10]. VoltageStorm® is a Cadence® tool that analyzes a chip's power distribution network for IR voltage drop and metal electromigration failure[11]. Following is the flow for using VoltageStorm for standard 2D designs modified for 3D designs. It is assumed this analysis will be used only for digital designs for which physical design steps are undertaken using CAD tools.

FLOW FOR 3D IR DROP ANALYSIS

1) Create 3D technology file:

The extraction tools need a technology file that contains predicted information of parasitic resistance and capacitance of an integrated circuit design. Fabrication process information is to be entered in an ASCII-format process description file (text file). [12] This file is then converted to a binary file using a Cadence® tool called Techgen®. For resistance extraction, the technology file contains resistance information on each interconnect layer and via. For capacitance extraction, the file contains three-dimensional interconnect models.

For 3D analysis, a 3D technology file is required. However, the process description file does not contain any substrate information but only on interconnect layers, vias and

dielectrics. Hence, a different approach is taken. Interconnect layers, vias and dielectric layers of all dies are made to appear to belong to just a single die, and intermediate substrates are defined as dielectrics with appropriate thickness. Figure 11 gives a pictorial representation of the process description file for a two-die design with face-to-back stacking and containing TSV for inter-die connections. This image was generated by ViewICT® tool from Cadence.

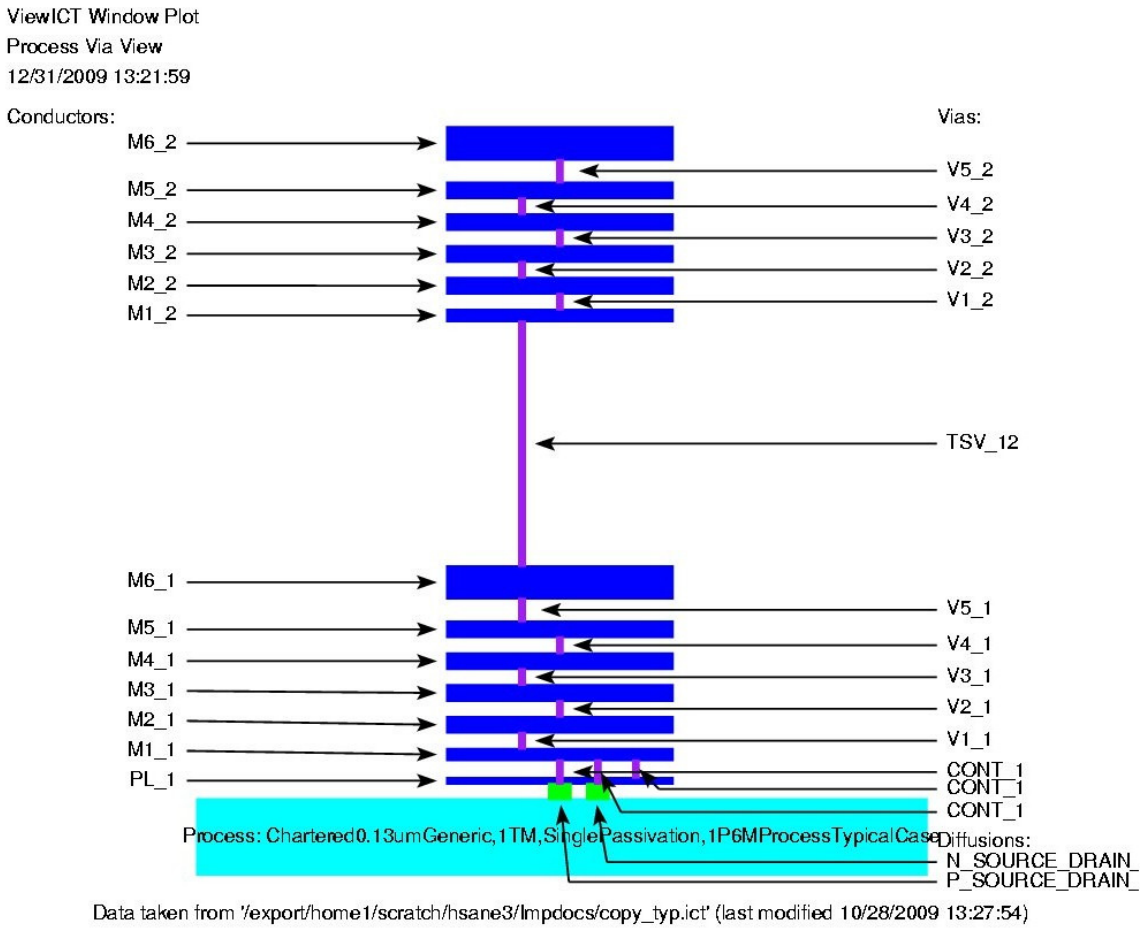


Figure 11: 3D process description file for face-to-back stacking

Through-silicon-vias form connections between metal interconnect layers of different dies and are represented using additional via definitions in the 3D process description file. Same interconnect metal layers and vias in different dies are defined separately. So METAL1 became M1_1 for die1, M1_2 for die2, and so on.

2) Modify LEF files for 3D:

LEF stands for Cadence® Library Exchange Format. LEF file defines the elements of an IC process technology and associated library of cell models [13]. The file is an ASCII representation of library data including layer, via, placement site type, and macro cell definitions.

For 3D analysis, 3D LEF files are required. Even if all cells and devices in the design remain 2D i.e. any device is completely on a single die only, the same device type may have multiple instances on different dies in a 3D IC. To differentiate between these multiple devices, multiple definitions of same library cell are required, one for each die in the design. Apart from this, new definitions for TSV arrays have to be added.

3) Modify layer mapping files for 3D:

DEF stands for Cadence® Design Exchange Format. DEF file contains design-specific information of a circuit, and is an ASCII representation of the design at any point during the layout [13].

To ensure Cadence® Encounter tool can match device in DEF file to the correct definition in LEF file, a mapping file is required to match naming conventions in LEF

and DEF file. For 3D design, this mapping file has to be modified to include additional layers of upper dies.

Instead of ASCII format, design can be represented in GDS format. In such cases, a mapping file is required to match naming conventions between LEF file and GDS file. However, GDS format gives layer numbers to standard layers in a design, such as interconnect metal layers and vias. Hence, it is not possible to name same layer for different dies in same file. As a result, a limitation of mapping only one die information to LEF file exists. So if custom blocks exist in more than one die, the analysis has to be performed once for every custom-block-containing-die. Analysis result for CAD tool designed section of the design will remain same, but block information will be available for only one die at a time.

4) To view power grid of macros, modify XTC file:

Separately designed blocks, especially full-custom blocks, can be included within another design by defining them as macro blocks in an LEF file. Generally full-custom blocks would be present in GDS format. Hence, noise analysis tools cannot look inside the block but considers it as a single unit. Hence, analysis results of entire block are averaged out to single number, and this reduces usefulness of the analysis. If the block contains a single point that exceeds design tolerance, its location will still remain a mystery. Hence, it is extremely advantageous if network inside the block is made visible to the tool for analysis.

XTC is a Cadence® tool used in the design flow. It prepares data for both extraction and analysis[14]. XTC tool has to be given instruction on how to perform extraction as per the design specifications. These commands are placed in a command file. Using this tool, supply grid for macro blocks can be extracted and made available for analysis. The XTC command file is used along with the LEF/GDSII mapping file. Since this mapping file has a limitation of providing information for one die at a time, the XTC command file too shows same limitation.

5) Create library:

Cadence® provides a cell library preprocessing tool LibGen, which is used to create and modify a cell library database of power-grid views[14]. LibGen takes LEF, GDSII or a combination of LEF and GDSII as input and creates a binary power-grid view of the cell content within the LibGen cell library database. A power-grid view is a model of a cell's or block's power grid that contains power port information providing the current loading to the power-grid networks.

6) Power related files:

VoltageStorm adds imaginary sources at points specified in a pad file, and performs noise analysis. This file has to be modified to run 3D analysis. The file actually contains (x, y) coordinated at which voltage sources are to be placed for analysis. Along with the coordinates, a physical location has to be specified in the form of interconnect metal layer name on which source is located. This naming has to be modified to suit 3D naming scheme used in LEF files.

To perform signoff noise analysis in 2D, a preliminary power analysis has to be run that produces a file containing predicted power dissipation in each instance present in the design based on inputs such as predicted toggling probability of nets or clock frequency. Similarly for 3D analysis, a similar file for each of the individual dies is required. These files have to be modified to suit naming convention for 3D as specified in LEF files, and then concatenated to a single file.

7) Modify DEF file for 3D:

As mentioned before, DEF file is an ASCII representation of the design as present in layout form. For 3D, certain changes need to be made to DEF file of each die before combining them into a single file. Component instances, via arrays as well as power and ground nets need to be renamed as per 3D naming convention. Pins and signal nets can be ignored as they are not required for IR drop analysis.

8) Run VoltageStorm

Once all files are ready, the VoltageStorm® tool is run as a stand-alone tool to obtain IR drop analysis results for the 3D IC design.

RESULTS FOR 3D NOISE ANALYSIS – 3DMAPS

The flow has been tested on different designs with success.

Two designs and analysis results are demonstrated here.

3DMAPS is a 64-core 3D-stacked memory-on-processor system. The processor contains 64 cores connected in a 2D mesh structure for communication. Each core has a 32-bit five-stage in-order VLIW pipeline and 4KB of dedicated local SRAM memory. The instruction memory contains 1.5KB of 64-bit two-instruction bundles. For this design, Chartered 130nm process with 6 metal layers has been used.

The two dies are stacked face-to-face.

Following are plots from IR drop analysis (VDD net) of a single core and single memory stacked one over the other. Vertical connections are assumed at all corners of the core to connect supply rails for core die to corresponding rail in memory die.

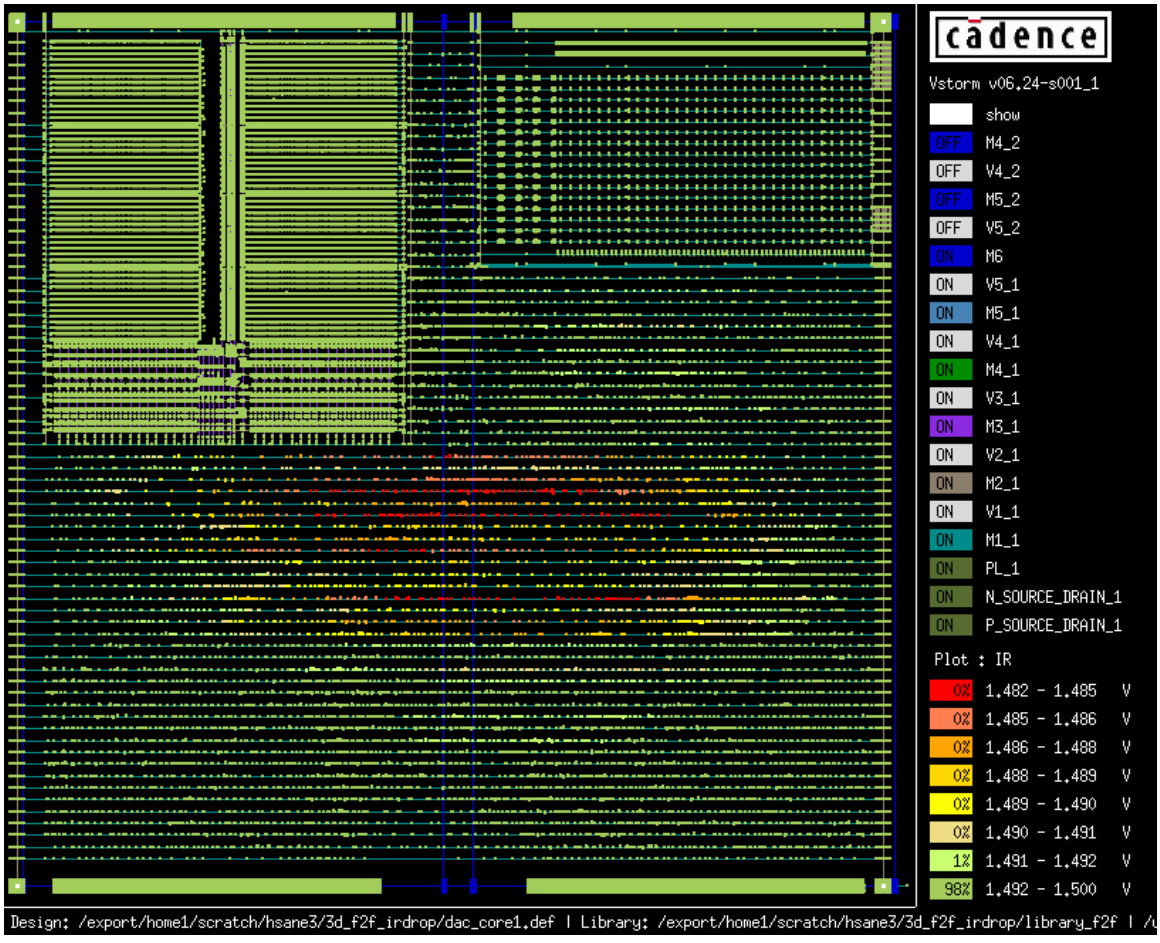


Figure 12: The 3DMAPS core die

Figure 12 shows VoltageStorm® analysis result of a single core. As seen on the right-side of the image, all metal layers are named as per their die number. Color on the plot indicates level of supply drop with respect to full rail value and allowable threshold value. Green indicates negligible drop, while red indicates threshold has been crossed. For this design, threshold of 15 mV on a supply of 1.5 volts was taken. Since this design is actually a part of the multi-core system, a reasonable lower value of threshold was assumed. Here, worst case drop is seen to be 18mV.

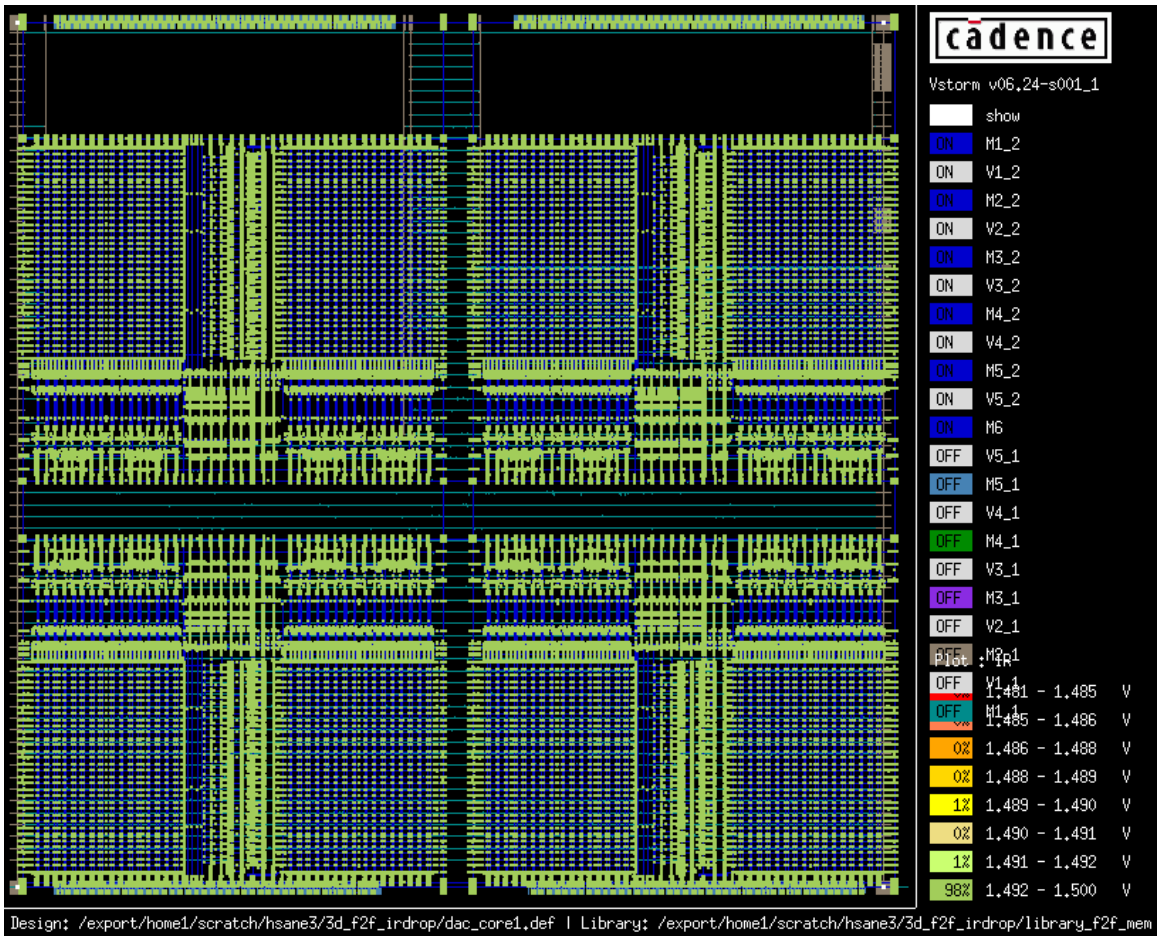


Figure 13: The 3DMAPS memory die

Similarly, the analysis results for memory die are shown here in figure 13. Worst case drop is seen to be 11mV. Memory die here contains macro blocks whose power-grid view has to be extracted. Hence internal power-grid of blocks lying in core die is not visible in this plot. This is one of the limitations of the described method for 3D IR drop analysis.

CADENCE Limitations for 3D IR drop analysis

Cadence® Encounter® has an established flow for performing IR drop analysis for 2D designs. As seen in previous section, it can be modified to handle 3D design containing TSVs. However, the modifications are only a quick-fix, as Cadence Encounter® has certain in-built features that restrict benefits of the above mentioned modifications.

- Cadence Encounter is built for handling 2D designs only. So, it can read and understand only one die at a time. This becomes a restriction when more than one die in 3D design have macro blocks whose power-grid needs to be extracted.
- VoltageStorm® does parasitic R extraction using Fire&Ice® tool from Cadence, and uses the extracted netlist to perform IR drop calculation. The R extractor has a limit on the number of layers it can handle. As seen in the modified flow, 3D is represented in process description file as single die having all the layers. However, due to the limitation on number of layers, this flow can handle only certain 3D designs.

CHAPTER 5

CONCLUSION AND FUTURE WORK

3D noise analysis has plenty of scope for innovation.

One direction is by using new 3D specific algorithms. The optimization algorithm is able to give results with huge savings in time. Adding a congestion constraint to the original problem can make the algorithm congestion aware and help reduce timing issues in the design. The algorithm in its current form is not very accurate, but is good enough for first order of approximation. And hence, it can allow a designer to continuously tweak the design and run multiple simulations without spending a large amount of time.

A second way forward is by expanding on existing tools for 2D to deal with 3D. The 3D IR drop analysis flow using Cadence® tools does have some limitations, but is a big step forward. If the limitations can be overcome, the current 3D flow can be made much more robust.

REFERENCES

1. Black, B., et al., *3D processing technology and its impact on iA32 microprocessors.* Proceedings of the IEEE International Conference on Computer Design, 2004: p. 316-318.

2. Kim, D.H., S. Mukhopadhyay, and S.K. Lim, *Through-Silicon-Via Aware Interconnect Prediction and Optimization for 3D Stacked ICs.* ACM/IEEE International Workshop on System Level Interconnect Prediction, 2009.

3. Zhou, P., K. Sridharan, and S.S. Sapatnekar, *Congestion-Aware Power Grid Optimization for 3D Circuits Using MIM and CMOS Decoupling Capacitors.* Proceedings of the 2009 Conference on Asia and South Pacific Design Automation, 2009.

4. Sapatnekar, S.S., *CAD for 3D circuits: Solutions and challenges.* Proc. VMIC, 2007.

5. Lee, Y.J., R. Goel, and S.K. Lim, *Multi-functional Interconnect Co-optimization for Fast and Reliable 3D Stacked ICs.* IEEE International Conference on Computer-Aided Design, 2009.

6. Healy, M. and S.K. Lim, *A Study of Stacking Limit and Scaling in 3D ICs: an Interconnect Perspective.* IEEE Electronic Components and Technology Conference, 2009: p. 1213-1220.

7. Huang, G., et al., *Power delivery for 3D chip stacks: physical modeling and design implication.* Proc. IEEE Conference on Electrical Performance of Electronic Packaging, 2007.

8. Synopsys, *Raphael.* http://www.synopsys.com/tools/tcad/interconnectsimulation/pages/raphael.aspx (Accessed January 04, 2010).

9. Ho, C.-W., A.E. Ruehli, and P.A. Brennan, *The Modified Nodal Approach To Network Analysis.* IEEE Trans. on Circuits and Systems, 1975.

10. Cadence, *Encounter User Guide.* 2008.

11. Cadence, *VoltageStorm Transistor-Level Rail Analysis User Guide.* 2007.

12. Cadence, *QRC Techgen Reference Manual.* 2008.

13. Cadence, *LEF/DEF Language Reference.* 2008.

14. Cadence, *VoltageStorm Data Preparation Manual.* 2007.